Rozdział 4 - Przetwarzanie danych - ćwiczenia

## Analiza statystyczna opisowa pliku *crim\_hom\_soff.tsv*

1. Wczytanie i transformacja pliku csv *crim\_hom\_soff.tsv*.

rm(list=ls())  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
Dane<-read.table("./Dane/crim\_hom\_soff.tsv",header = T,  
 dec=",",sep = "\t",encoding = "UTF-8",stringsAsFactors = F)  
colnames(Dane)<-sub("X","",colnames(Dane))  
Kol1<-Dane[1]  
Dane <- Dane %>% select(-1)  
Dane<- Dane %>% mutate\_if(is.character,as.numeric)

## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia  
  
## Warning: pojawiły się wartości NA na skutek przekształcenia

Kol1<-Kol1 %>% separate(1,c("Przestępstwo","Typ","Płeć","Jednostka","Kraj"),sep=",",convert = T)  
Kol1<-Kol1 %>% mutate\_if(is.character,as.factor)  
Dane<-bind\_cols(Kol1,Dane)

1. Przeksztalcenie na postać wąską

Dane2<-gather(Dane,Rok,Wartość,'2008':'2017')  
Dane2<-Dane2 %>% mutate\_at("Rok",as.numeric)  
Dane2<- Dane2 %>% filter(Jednostka=="P\_HTHAB",Typ=="PER\_VICT",Przestępstwo=="ICCS03011",Płeć!="T")  
Dane2<- Dane2 %>% filter(complete.cases(Dane2))

1. Statystyki

tabela1<-Dane2 %>% group\_by(Kraj) %>% summarise(suma=sum(Wartość)) %>%  
 arrange(desc(suma)) %>% mutate(miejsce=rank(rev(suma))) %>%  
 filter(miejsce<=6 | Kraj=="PL")   
  
tabela2<-Dane %>% group\_by(Kraj) %>%   
filter(Jednostka=="P\_HTHAB",Typ=="PER\_VICT",Przestępstwo=="ICCS03011",  
 Płeć %in% c("F","M")) %>%   
 select(c(Kraj,'2008':'2017')) %>% summarise\_each(sum)  
średnia<-round(sum(Dane2$Wartość) /nlevels(Dane2$Kraj),2)  
zestawienie <-tabela1 %>% left\_join(tabela2,by="Kraj")  
library(knitr)  
kable (zestawienie)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kraj | suma | miejsce | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
| SE | 1314.15 | 1 | 117.97 | 127.72 | 127.12 | 138.29 | 133.02 | 125.66 | 138.66 | 121.40 | 136.46 | 147.85 |
| UKC-L | 937.50 | 2 | 47.20 | 53.95 | 56.48 | 56.57 | 57.37 | 57.00 | 112.84 | 136.59 | 156.92 | 202.58 |
| UKN | 656.64 | 3 | 40.84 | 46.45 | 54.35 | 59.62 | 54.52 | 58.29 | 73.81 | 80.53 | 87.92 | 100.31 |
| NO | 509.85 | 4 | 28.86 | 29.65 | 28.18 | 31.22 | 30.46 | 61.44 | 61.68 | 71.67 | 83.11 | 83.58 |
| UKM | 405.88 | 5 | NA | NA | 9.38 | 30.22 | 36.54 | 61.81 | 65.59 | 61.58 | 63.72 | 77.04 |
| DK | 347.93 | 6 | 35.27 | 31.89 | 32.51 | 29.98 | 29.08 | 27.93 | 27.81 | 32.90 | 45.15 | 55.41 |
| PL | 7.76 | 31 | NA | NA | NA | NA | NA | NA | NA | NA | 4.06 | 3.70 |

1. Wykres

library(ggplot2)  
  
ggplot() +   
 geom\_bar(data=zestawienie,mapping=aes(y=suma,x=reorder(Kraj,-suma),fill=Kraj),stat="identity") +  
 theme(legend.position = "") + xlab("Kraj") + geom\_hline(yintercept = średnia)+  
 ylab("Ilość gwałtow na 100tys. mieskanców")+  
 geom\_text(aes(x=c(3,7),y=c(średnia+100,średnia+100),label=c("Średnia ilość gwałtów w Europie liczona na każdy kraj",paste(średnia))))+  
 geom\_text(data=zestawienie,mapping = aes(x=Kraj,y=-100,label=paste("miejsce",miejsce,sep=" ")))

